Online Social Regulation

Roos, Carla A.; Koudenburg, Namkje; Postmes, Tom

Published in:
Journal of Computer-Mediated Communication

DOI:
10.1093/jcmc/zmaa011

Publication date:
2020

Online Social Regulation: When Everyday Diplomatic Skills for Harmonious Disagreement Break Down

Carla A. Roos , Namkje Koudenburg, & Tom Postmes

Department of Social Psychology, University of Groningen, 9712 TS Groningen, The Netherlands

In group discussions, people rely on everyday diplomatic skills to socially regulate the interaction, maintain harmony, and avoid escalation. This article compares social regulation in online and face-to-face (FtF) groups. It studies the micro-dynamics of online social interactions in response to disagreements. Thirty-two triads discussed, in a repeated measures design, controversial topics via text-based online chat and FtF. The fourth group member was a confederate who voiced a deviant (right-wing) opinion. Results show that online interactions were less responsive and less ambiguous compared with FtF discussions. This affected participants' social attributions: they felt their interaction partners ignored them and displayed disinhibited behavior. This also had relational consequences: participants experienced polarization and less solidarity. These results offer a new perspective on the process of online polarization: this might not be due to changes in individual psychology (e.g., disinhibition), but to misattributions of online behavior.

Keywords: Social Regulation, Online Discussions, Disinhibition, Polarization, Diplomacy

doi:10.1093/jcmc/zmaa011

Many people believe that social media undermine civility (Weber, Powell , KRC Research, 2013). Indeed, a large scientific literature documents effects of social media on, among others, polarization, sectarianism, and social exclusion (e.g., Anderson, Yeo, Brossard, Scheufele, & Xenos, 2018; Coe, Kenski, & Rains, 2014). There are many ideas about the psychological underpinnings of such phenomena. Some suggest that online, people experience fewer social constraints which fosters disinhibition (e.g., Suler, 2004; Voggeser, Singh, & Göritz, 2018). Others have pointed out that to the contrary, people are often *more* focused on social norms and on each other online (e.g., Lea, O'Shea, Fung, & Spears, 1992; Spears, Postmes, Lea, & Watt, 2001; Walther, 1992).

We contribute to the existing literature by offering a new perspective. Instead of assuming that the medium changes individuals' psychology (e.g., anonymity leading to deindividuation), which in

Corresponding author: Carla Anne Roos; e-mail: c.a.roos@rug.nl
Editorial Record: First manuscript received on 20 November 2019; Revisions received on 17 April 2020; Accepted by Nicole Kraemer on 10 June 2020; Final manuscript received on 17 June 2020

turn changes behavior (e.g., disinhibition), we propose that the medium directly changes behavior, which affects social perceptions and thereby changes relationships. We compare text-based online and face-to-face (FtF) group discussions on politically controversial topics in which contentious opinions are voiced. What diplomatic skills do group members use to socially regulate contentiousness, and how does this affect their perceptions of each other and their social relationships? We propose that socially regulating online conversations is more difficult for two reasons: (a) online interactions are less responsive due to a lack of synchronicity, and (b) online utterances are more explicit and unambiguous due to a limited ability to convey subtle social cues. We expect that interaction partners will misattribute this unresponsiveness and excessive clarity to a lack of social concern, polarization, and conflict.

This research extends the literature in two ways. First, applying the findings of the pragmatics literature to the differences between text-based online and FtF discussions leads us to the provocative prediction that online discussions are relatively unambiguous and that ambiguity can be a good thing. Moreover, we incorporate online sender–receiver dynamics and social (mis)attributions. Instead of assuming that people are less socially concerned online, or that the medium psychologically transforms them in another way, we propose that the intrinsic characteristics of online interaction can contribute to a negative sender–receiver dynamic: the relative unresponsiveness and clarity of senders' messages may lead receivers to feel ignored and rejected, which in turn affects perceived polarization and solidarity.

Everyday diplomacy and social regulation

Our point of departure is empirical research that suggests that when people encounter strong differences of opinion FtF, it is quite uncommon to take an explicit stance. Instead, people signal disagreement in subtle and implicit ways (e.g., a frown, a short silence) coupled with considerable ambiguity in message content (Bavelas, Black, Chovil, & Mullet, 1990; Brennan & Clark, 1996; Reid, Keerie, & Palomares, 2003). By using such techniques, communicators diplomatically signal their disagreement: they are still able to maintain harmony because they implicitly sanction others for violating social norms (Koudenburg, Postmes, & Gordijn, 2017). Because they are more common than explicit reprimands and sanctions, such everyday diplomatic skills appear to fulfill an important role in maintaining harmony in groups.

Disagreement in conversation is often avoided because it can harm social relationships (Brown & Levinson, 1987; Pomerantz, 1984). In an attempt to maintain harmony whilst navigating disagreements, people engage in social regulation. In line with Social Information Processing theory (Walther, 1992), we assume that most people seek to prevent conflict and maintain harmonious social relationships also online. Indeed, whereas in some online contexts incivility is frequent (e.g., Coe et al., 2014), in many online discussions incivility is rare (Papacharissi, 2004). But online interactions can sometimes be fractious, for example in situations where politically controversial topics are discussed on Internet fora, where there is a high a priori likelihood of disagreement coupled with an absence of established relationships. In our digitizing world, such discussions are increasingly common forms of "doing" politics. This implies that social regulation remains very important online. In order to develop a better understanding of online social regulation, it is informative to first look at what we know about the ways in which people handle disagreements in FtF conversations. We focus on two everyday diplomatic skills known from the pragmatics literature that we expect to be less available online: responsiveness and ambiguity.

Social regulation FtF: Responsiveness

Research shows that interaction partners rely on the flow of a conversation to gauge the status of their social relationship. Indeed, responsiveness, here defined as the degree to which interaction partners provide instant feedback to each other, fulfills an important social function. People frequently interject words ("yes"), vocalizations ("hmm"), or head nods during another speaker's turn or at the start of their own turn (Beňuš, Gravano, & Hirschberg, 2011). This signals attentiveness but also has a wider significance by conveying that one is "with" the other speaker in the sense of adopting a socially shared understanding and consensus (Clark, 1996; Koudenburg et al., 2017). The receiver interprets these as signals of interpersonal interest and attraction as well as social harmony (Davis & Perkowitz, 1979; Reis & Clark, 2013). Conversely, people infer misunderstanding, dissent, and social rejection from silences and other interruptions in the flow of interactions, even if those are clearly due to factors beyond their interaction partner's control, such as a delay in the communication channel (Koudenburg, Postmes, & Gordijn, 2013).

In sum, responsiveness during conversation promotes harmony. This applies to all everyday interactions, but might be especially important in contentious discussions. Ambiguity is a diplomatic skill that is more exclusively associated with contention.

Social regulation FtF: Ambiguity

People usually try to communicate clearly and directly (Grice, 1975), but sometimes this may damage social relationships. In such cases people tend to communicate more indirectly, ambiguously, or evasively (Bavelas et al., 1990; Brown & Levinson, 1987). Indeed, research shows that people pre-empt conflict by ambiguating their message rather than expressing their disagreement clearly (Pomerantz, 1984). In everyday conversation, people can ambiguate with disclaimers (e.g., "I do not know for sure"), hedges (e.g., "maybe," "sort of"), and vocalizations that express doubt (e.g., a drawn out "hmmm") or tentativeness (e.g., "uhm," Brennan & Clark, 1996; Koudenburg et al., 2017; Reid et al., 2003). Whilst excessive vagueness can irritate, a modest amount conveys thoughtfulness, modesty, and a consideration for others' views and positions (Geddes, 1992). Substantively, ambiguity makes disagreement less likely and creates scope for interaction partners to assume consensus (i.e., social projection, Krueger, 1998). Ambiguity during a contentious discussion can therefore help to avoid (the escalation of) disagreement and cement social relationships.

In sum, responsiveness and ambiguity are everyday diplomatic skills used to maneuver through disagreements whilst preserving social harmony. Both factors are more about the *style* of expression than about content. How do these skills fare online?

Social regulation online

Online discussion differs in many respects from FtF discussion. Two key characteristics of text-based online media are its a- or semi-synchronicity and its relative lack of subtle social cues (e.g., Alberici & Milesi, 2018; Kiesler & Sproull, 1992; Suler, 2004). To the extent that these characteristics interfere with the techniques used for social regulation in FtF conversations, they might make maintaining harmony more challenging.

One difference is that the relative lack of synchronicity in online discussions hinders the instant feedback that is common in FtF social regulation (e.g., Daft & Lengel, 1986; Suler, 2004). Based on this, we predict that, compared to FtF conversations, text-based online discussions will be relatively unresponsive (Hypothesis 1a). Moreover, the difficulty of conveying subtle social cues makes it more difficult for online users to use the ambiguation techniques that they rely on in FtF social regulation. Indeed, in text-based online communication, subtle cues tend to be replaced by more explicit verbal

cues (Walther, Loh, & Granka, 2005). We therefore propose that participants are prone to stating their views more clearly online (Hypothesis 1b). We have no reason to expect that participants will express any more (or less) disagreement online (Hypothesis 1c)., we

Taking this first set of hypotheses together, we expect text-based online interactions to be relatively unresponsive and clear. Because of the importance of responsiveness and ambiguity for the social regulation of FtF discussions, we expect that online interaction partners might be more prone to misinterpret each other's intentions and motivations. People typically attribute each other's behavior to internal causes (e.g., predispositions, personality) rather than recognizing situational explanations beyond the other's control (fundamental attribution error, Jones & Harris, 1967). Accordingly, online interaction partners might not recognize that the cause for the reduced responsiveness and enhanced clarity of language lies in the restrictions the medium imposes on behavior but rather attribute this to the sender's self-absorbedness and/or a-sociality. Specifically, participants might conclude that their interaction partners are ignoring them (Hypothesis 2a) and showing disinhibited behavior (Hypothesis 2b) more in text-based online than in FtF discussions.

Finally, these online micro-dynamics might lead people to the conclusion that there must be a problem at the social level: their partners might disagree with or reject them (see also Koudenburg et al., 2013, 2017). Consequently, participants might experience less consensus (i.e., more polarization, Hypothesis 3a) and less solidarity (Hypothesis 3b) within their group when they discuss controversial topics online compared to FtF.

When modeling these effects (see Figure 1), we expect that the relative *lack of responsiveness* together with the *increased clarity*, predicts misattributions of ignoring and disinhibited behavior (Hypothesis 4a), and the experience of increased polarization and reduced solidarity (Hypothesis 4b) among online interaction partners. In other words, regardless of their self-reported levels of disinhibition and actual polarization, participants will *experience* more disinhibition and polarization online due to the relatively unresponsive and clear style of expression. We thus expect a mismatch between actual and perceived disinhibition and polarization that is driven by expression style.

A recent study (Roos, Postmes, & Koudenburg, 2020) found initial evidence of a lack of responsiveness and abundance of clarity in text-based online chats and suggested this could undermine social harmony. However, this was a more explorative study that did not zoom in on social regulation because there was little need for it: participants tended to agree. Moreover, because of the explorative nature of the study, the authors could not examine underlying processes. In this follow-up study, we therefore introduced disagreement experimentally and examined social misattributions as a possible underlying process.

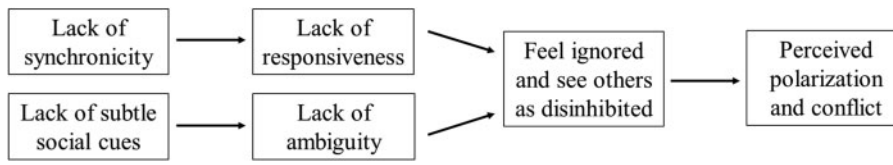

Figure 1 Consequences of the breakdown of everyday diplomatic skills in text-based online discussions: A visual representation of the proposed model.

Media richness

It is useful to compare the hypotheses to the broader literature on ambiguity in online communication. Most literature in this domain builds on the ideas set out in Media Richness Theory (Daft & Lengel, 1986). It has been assumed that communication media that are relatively rich in social cues and allow for immediate feedback (e.g., FtF) tend to reduce ambiguity, while lean media that lack many social cues and hinder immediate feedback (e.g., text-based chat) tend to increase ambiguity (e.g., Runions, Shapka, Dooley, & Modecki, 2013). Further, ambiguity is assumed to be problematic because it feeds misunderstanding (Edwards, Bybee, Frost, Harvey, & Navarro, 2017). In contrast to this literature, we propose that the style of text-based online interaction is relatively clear and *un*ambiguous. Inspired by the pragmatics literature, we further propose that ambiguity can be a virtue when it comes to regulating social relationships in contentious conversation.

Disinhibition

Our predictions also extend the online disinhibition literature. This literature assumes that because the online environment lacks the subtle social signals and instant feedback of FtF interactions, people are liberated from social constraints and unconcerned about others' evaluations (e.g., Suler, 2004). This will result in a disregard for social norms and behavioral disinhibition (a process similar to deindividuation, see Postmes & Spears, 1998). This can take a benign or a toxic form. The latter has received most research attention and is most relevant in the context of this paper. Toxic disinhibition involves acts like name-calling, rude language, threats, and other forms of hostile and aggressive behavior (often referred to as "flaming," Lapidot-Lefler & Barak, 2012).

Evidence for the online disinhibition effect is inconsistent, however (Clark-Gordon, Bowman, Goodboy, & Wright, 2019; Lea et al., 1992; Spears et al., 2001). For example, people appear to be very susceptible to social influence and able to form intimate social relationships online (Spears et al., 2001; Walther, 1992). Despite these findings, the idea that being online disinhibits remains widespread in the online communication literature (e.g., Casale, Fivaranti, & Caplan, 2015; Lapidot-Lefler & Barak, 2012; Voggeser et al., 2018) and beyond (e.g., Terry & Cain, 2016).

As outlined above, we sidestep this issue by focusing on behavior and the way interaction partners socially interpret and attribute this. We thus focus on the psychology of the receiver rather than the sender. We propose that any disinhibition observed may not (just) be caused by psychological changes leading to the adoption of more radical stances, but result from the restrictions that the medium imposes on people's expressions: unresponsiveness is interpreted as a sign of disinterest and clarity of expression is seen as outspokenness and this results in the *impression* of disinhibition. Thus, online disinhibition may be *perceived* rather than *enacted* or *intended*.

Research overview

In order to test the hypotheses, we designed a multilevel repeated measures experiment in which small groups of unacquainted Dutch student participants discussed about controversial topics via both a text-based online chat and FtF. One of the group members was a confederate who introduced disagreement by voicing a right-wing opinion amidst the mostly left-wing participants. We content coded participants' immediate reactions to assess the behavioral effects (Hypothesis 1). To assess participants' conversational experiences (Hypotheses 2 and 3), participants filled out a self-report questionnaire after each discussion. The connection between the behavioral effects and the conversational experiences proposed in Hypothesis 4 was tested by correlating the content coding with the

questionnaire scores. Lastly, actual polarization in response to the confederate's diverging standpoint was explored by assessing whether and how participants' private attitudes changed from pre- to post discussion, both with respect to the confederate's position and internally within the group.

Method

Pilot study

Fifteen first-year Dutch psychology students rated 30 opinion statements that were selected because they differed along the current left- and right-wing divide in Dutch politics, and were deemed relevant and interesting for students. As stimulus material for the main study, we selected four statements that students: (a) rated as clear, (b) had a clear opinion about, (c) had a similar opinion about, (d) expected to agree on with other students, and (e) found interesting. Of these four statements, two were leftist: "Dutch businesses that support child labor should be punished heavily" and "People whose asylum claims have been dismissed are entitled to food, drink, and shelter," and two were rightist: "It is natural that there are few women in the top of business and government" and "No more mosques should be built in the Netherlands."

Research design

The main study had a multilevel repeated measures design: each group of participants took part in both the FtF and the text-based online condition. In each condition, groups discussed two statements consecutively. The confederate introduced disagreement in each second conversation. Only these conversations were analyzed. The allocation of groups to combinations of condition and discussion statement orders was based on a Graeco-Latin square with eight unique cells (Walker & Lev, 1953). This square was constructed by combining condition orders (two conditions) with statement orders (four statements) in such a way that each condition and each statement occurs in each cell, but in a different order so that each combination occurs only once in the entire square. This design enables us to exclude order effects for condition and statement.

Power and sample size

Taking into account our design, we calculated the required sample size by means of a simulation-based power analysis for mixed models in lme4 in R (Bolker, 2014). This simulation was based on the results of a previous study with a similar design (Roos, Postmes, & Koudenburg, 2020). Thirty-two groups of three participants were required to achieve a power of .85 for a two-sided test of the smallest effect obtained in the previous study (an item measuring disagreement, part of the perceived consensus scale, which had a Cohen's *d* of .24). Since this was the smallest effect and power for the full consensus scale approached 1, we assumed this to be sufficient for the detection of effects.

Participant characteristics

Participants[1] were 96 native Dutch (M_{age}=20.20, SD_{age}= 3.20; 85.42% female) students who participated for partial course credit or monetary compensation. Most participants did not know their group members before the experiment started (84.37%). As expected, the political orientation of the sample was skewed to the right: 57.29% left-wing, 33.33% moderate, and 9.38% right-wing.

Procedure and apparatus

Participants were invited into the lab in triads. Upon their arrival, in order to avoid interaction, participants were immediately separated in cubicles with a computer. The experimenter gave each of them individually a short introduction to the chat software (Google Hangouts). She made clear that they were allowed to use emoji.

Groups then discussed four statements: two via text-based online chat and two FtF. They engaged in the text-based online chat (in which they were pseudonymized) in their individual cubicles. The FtF discussions took place in an adjacent room where participants were seated in a circle. Rotated according to the Graeco-Latin square design, right-wing and left-wing formulated discussion statements were alternated. Participants were allowed to discuss each statement for up to 10 minutes but could collectively decide to end their discussion before.

One of the group members was a confederate, whose demographics matched those of the participant sample: one of two Dutch female psychology students. In each first conversation in each condition, the confederate did not have any pre-set text and was instructed to agree with what was being said. These were meant as get-acquainted conversations and were not analyzed. However, in the second discussion in each condition the confederate expressed a dissenting viewpoint. Her first two sentences were scripted. They were designed to sound like natural and credible right-wing positions, e.g., sentence one: "Hmm, I don't know. Evolutionary, women are simply not used to lead groups, they are used to care" and sentence two: "Well, only women become pregnant and work more part-time, which is, of course, not really possible in top positions." The confederate spoke or wrote the first scripted sentence after at least two participants expressed their opinion and were in (mostly left-wing) agreement. She introduced the second sentence at a natural moment later in the conversation. To allow for natural conversation, the remainder of the discussion was unscripted, but the confederate was instructed to stick to her position.

Chat interactions were stored and FtF conversations were audio-recorded. After each discussion in each medium (four times), participants filled out a self-report questionnaire on their computers. At the end of the entire experiment, participants provided demographic data[2] and were debriefed. The presence of confederates was disclosed to participants in a follow-up email after the data collection was finished.

Dependent measures

Content coding

We adapted the coding scheme of Roos, Postmes, & Koudenburg (2020), which will be detailed below. One Dutch student-assistant, unaware of the research hypotheses, was trained as coder. The first author acted as the second coder. Both independently coded all conversations in randomized order. We coded the untranscribed audio-recordings of the FtF interactions in order to retain more of these conversations' style (e.g., intonation). To assess the inter-rater reliability of the ordinal (mostly Likert scale) variables, we calculated two-way absolute agreement average measures intra-class correlation coefficients (Hallgren, 2012).

We coded all the speaking turns of participants between the confederate's first and third statement, which represents the standardized part of the conversations. This focused the coding on the social regulation: the group's immediate reactions to a dissenter. We defined turns, based on Beňuš et al. (2011), as expressions that were successful in taking the floor, did not entirely overlap with another utterance, and held more content than only laughing or humming. We averaged the turn-by-turn ratings to obtain one score per conversation per group.

Hypothesis 1 concerns behavioral effects. In order to test Hypothesis 1a, we measured responsiveness by indicating for each turn whether it connected to the turn directly preceding it (1 = *No*, 2 = *A bit*, 3 = *Yes*; ICC[3] = .80, 95% CI [.77, .84]). When a turn started with a connecting word (e.g., "yes," "no," "but") and contained a reaction to the preceding turn, we coded it as responsive. When a connecting word was missing but the previous speaker was acknowledged, we coded it as a bit responsive. When the previous speaker's turn was ignored, this was classified as unresponsive. We tested Hypothesis 1b by rating the clarity of each turn (1 = *Very ambiguous*, 2 = *Ambiguous*, 3 = *Neutral*, 4 = *Clear*, 5 = *Very clear*; ICC = .66, 95% CI [.60, .72]). Generally, the more and the stronger the expressed ambivalence, disclaimers, and hedges (e.g., "I don't know for sure," "as far as I know," "sort of"), the more ambiguous a statement was considered (see also Reid et al., 2003). When participants presented their opinion as a fact, this was rated as very clear. Neutral was used for statements that were neither clear nor vague, which were often questions. To test Hypothesis 1c, we indicated for each turn to what extent it was in agreement (affirmative comment) or disagreement (criticizing comment) with the preceding turn that it seemed to refer to (-1 = *disagree*, 0 = *Neutral*, 1 = *agree*; ICC = .84, 95% CI [.81, .86]). When participants referred to the stimulus statement or their own prior remark, this was coded as neutral.[4] Because inter-rater reliabilities were adequate (Cicchetti, 1994), we performed the analyses on the means of the coders' ratings.

Questionnaire

After each discussion, participants filled out the same short questionnaire on their computers.[5] Hypothesis 2 concerns social misattributions. To assess whether participants felt ignored (Hypothesis 2a), we constructed a three-item scale with good reliability. Participants indicated how often they observed the following behaviors in their group: listening to each other, ignoring each other (R), and cross talking (1 = *Never* to 5 = *Continuously*; ω[6] = .78, 95% CI [.70, .83]). To test whether participants perceived inhibition (Hypothesis 2b), we included two items: "During this conversation I found the other participants polite" and "The other participants thought carefully about how they expressed themselves in this conversation" (1 = *Completely disagree* to 5 = *Completely agree;* ω = .62, 95% CI [.51, .72]). To get an indication of how perceived inhibition compares with actual inhibition, we included a self-rating equivalent to the last item.

Hypothesis 3 concerns relational consequences. We tested Hypothesis 3a by including a seven-item perceived consensus scale based on Koudenburg et al. (2013),[7] for example: "During this conversation I felt that we understood each other" and "During this conversation it became clear that we disagreed completely" (1 = *Completely disagree* to 5 = *Completely agree*; ω = .91, 95% CI [.88, .92]). To test Hypothesis 3b, we measured perceived solidarity with eight items: seven items adapted from Koudenburg, Postmes, Gordijn, and Van Mourik Broekman (2015),[8] for example "During this conversation I identified with the other participants," plus the following self-devised item "During this conversation the mutual relations in the group were good" (1 = *Completely disagree* to 5 = *Completely agree;* ω = .87, 95% CI [.82, .90]).

Lastly, participants' actual polarization in response to the confederate's diverging standpoint was assessed by looking at private attitude change from pre- to post-discussion. For this, participants indicated their agreement with the discussed statements (1 = *Completely disagree* to 7 = *Completely agree*) at the start and the end of the experiment.[9]

Statistical analyses

As participants were part of a group and were measured two times (in both conditions), the statistical analysis had to take into account these two sources of non-independence of observations. The

intraclass correlations (ranging from .03 to .39 at the group level and from .23 to .47 at the participant level) of the dependent variables suggested that scores were indeed clustered within groups and/or participants (Bliese, 2000). We therefore performed multilevel repeated measures regression analyses with condition (repeated measures; level 1), nested in participants (level 2), nested in groups (level 3). We analyzed the data with the lmer function in the R package lme4 (version 1.1-21, Bates et al., 2019). For all dependent variables, we compared the fit of the multilevel repeated measures model that included only the random effect(s) of participant and/or group with the equivalent model that added communication medium as fixed effect predictor. We used the emmeans package (version 1.4.1, Lenth, Singmann, Love, Buerkner, & Herve, 2019) to estimate condition means and confidence intervals. Because in all analyses the main and interaction effects involving condition order or discussion statement did not significantly improve the model fits, we did not include these factors in our analyses.

Results

The results will be presented in three sections. The first section covers the exploration of private attitude dynamics. The second section describes the content analysis showing the behavioral differences between conditions (Hypothesis 1). The third section focuses on social attributions and relational outcomes (Hypotheses 2 and 3). Lastly, section four tests the model by connecting behavior to attributions and outcomes (Hypothesis 4).

Private attitude dynamics

In order to explore the effect of communication medium condition on actual polarization in private attitudes, we calculated the degree to which participants' attitudes shifted towards the standpoint defended by the confederate. We recoded the attitude scores so that higher scores represented more agreement with the confederate. We subsequently entered the post-discussion attitude as dependent variable in a multilevel repeated measures model with communication medium as fixed effect predictor nested in participants nested in groups and pre-discussion attitude as covariate. There was no significant improvement in model fit compared to the model that only contained the random intercepts and the covariate ($\chi^2(1) = .67$, $p = .413$). This means that there was no effect of condition on the degree of attitude shift vis-à-vis the confederate's standpoint.

In a similar vein, we looked at the effect of condition on the degree of variability in private attitudes within groups. We calculated the within-group post-discussion attitude variability as the absolute difference between participants' post-discussion attitudes and their group means. The fit of the model did not improve after including condition as fixed effect predictor compared to the empty model with only the random intercepts and the pre-discussion attitude variability as covariate ($\chi 2(1) = 3.40$, $p = .065$). In sum, these results show that the medium had no significant effect on the degree of polarization of private attitudes on the discussed topic, either with respect to the confederate's position or internally within the group.

Behavioral effects

Discussion content was analyzed in multilevel repeated measures models with condition as fixed effect predictor nested in the random effect of groups. First, it is important to note that the coders did not observe any instances of aggressive or hostile language, swearing, derogatory names, etc.

(Lapidot-Lefler & Barak, 2012). This means that we found no evidence for toxic disinhibition in this study, despite the potentially polarizing position taken by the confederate.

As can be seen in Table 1, there were substantial between-condition differences for responsiveness and clarity. In text-based online chats, the first speaking turns after encountering a firmly worded disagreement were both less responsive and clearer compared with the FtF discussions ($d = 3.14$ and $d = -2.05$, respectively; strong effects, see Cohen, 1992). These findings support Hypotheses 1a and 1b. In line with Hypothesis 1c, there was no effect of communication medium on the amount of expressed (dis)agreement ($d = 0.34$).

The following two quotes (translated into English) taken from a FtF and an online discussion about the statement "It is natural that there are few women in the top of business and government" illustrate the distinction between ambiguity and clarity:

> *So then I do not really agree with it, but I do agree a little bit actually, because, I mean, it is the case that women always, well, women have children and ehm, yes, you are nevertheless in any case a bit more together with your family than a man, say, and a bit less, maybe, interested in a good job, so in that respect a little bit, but not as strong as it is now, I think. (FtF, Group 26)*
> *I don't agree with it. I think that the assertion that it is naturally determined is not correct. (Chat, Group 13)*

As is illustrated by this example, FtF comments are often dotted with hedges ("maybe"), hesitations ("ehm," "yes"), and sometimes explicit ambivalence ("I do not really agree with it, but I do agree a little bit"). Online chat comments, by contrast, are often more succinct, contain few ambiguating cues, and are therefore clear and explicit.

Social attributions and relational outcomes

We analyzed the questionnaire data in multilevel repeated measures models with communication medium as fixed effect predictor nested in participants nested in groups. The results are presented in Table 2. Online, participants experienced significantly more ignoring by and less inhibition of their group members than when discussing with them FtF ($d = -0.77$ and $d = 0.36$, respectively). This is consistent with Hypotheses 2a and 2b. Notably, participants did not consider *themselves* to be less inhibited online ($d = .02$). This means that participants thought they were equally thoughtful in constructing their messages but to *others* came across as if they were less socially considerate online. Results are also consistent with Hypotheses 3a and 3b: online, participants experienced less consensus and less solidarity ($d = .43$ and $d = .46$, respectively) in their group compared to when they discussed FtF. In sum, the pattern of results for social attributions and relational consequences (as well as actual disagreement and polarization) is consistent with expectations.

Connecting behavioral effects to social attributions and relational outcomes

Table 3 shows the group level repeated measures correlations between the content coding and the questionnaire data.[10] The sizable magnitude of some of these correlations (Cohen, 1988) indicates how influential the direct aftermath of encountering disagreement (a relatively small part of the discussion) was for participants' perceptions of the *whole* interaction and social relationships within the group *in general*.

First, discussion content did not relate to perceived inhibition of the self (r ranging between $|.00|$ and $|.04|$). Second, (dis)agreement in discussion content did not relate to participants' experiences to a significant degree (r ranging from $|.16|$ to $|.33|$). Although we did not formulate hypotheses around

Table 1 For each code, the test results of the difference between conditions (chi-square test and effect size) and the means with 95% confidence intervals per condition

	$\chi 2(1)$	FtF *M* [95% CI]	Chat *M* [95% CI]	d^a
Responsiveness	81.09***	2.52 [2.42, 2.62]	1.63 [1.53, 1.73]	3.14
Clarity	47.00***	3.02 [2.86, 3.17]	3.89 [3.74, 4.05]	−2.05
Expressed Agreement	2.35ns	-0.05 [-0.21, 0.11]	-0.21 [-0.37, -0.05]	0.34

Note. ns $p > 0.05$ * $p < 0.05$,
** $p < 0.01$, *** $p < 0.001$.
[a]Cohen's d was calculated by subtracting the chat estimates from the FtF estimates and dividing this by the total standard deviation of the full model (Cohen, 1988). This is a rather conservative estimate for effect sizes in repeated measures designs.

Table 2 For each questionnaire variable, the test results of the difference between conditions (chi-square test and effect size) and the means with 95% confidence intervals per condition

	$\chi 2(1)$	FtF *M* [95%CI]	Chat *M* [95%CI]	d^a
Perceived Ignoring	36.74***	1.68 [1.54, 1.82]	2.20 [2.06, 2.34]	−0.77
Perceived Inhibition others	12.27***	4.08 [3.92, 4.24]	3.83 [3.67, 3.99]	0.36
Perceived Inhibition self	0.02ns	4.12 [3.98, 4.25]	4.10 [3.97, 4.24]	0.02
Perceived Consensus	17.77***	2.91 [2.68, 3.14]	2.55 [2.33, 2.78]	0.43
Perceived Solidarity	17.43***	3.63 [3.48, 3.77]	3.33 [3.19, 3.47]	0.46

Note. ns $p > 0.05$,
* $p < 0.05$,
** $p < 0.01$,
*** $p < 0.001$.
[a]Cohen's d was calculated by subtracting the chat estimates from the FtF estimates and dividing this by the total standard deviation of the full model (Cohen, 1988). This is a rather conservative estimate for effect sizes in repeated measures designs.

these correlations, their non-significance is in line with our expectations: no condition effects on disinhibition of the self and expressed disagreement.

The other correlations provide strong support for Hypothesis 4. The correlations between responsiveness and participants' experiences were particularly strong (r ranging from |.48| to |.76|). More

Table 3 Repeated measures correlations between discussion content coding (columns) and participants' perceptions as assessed by self-reports (rows) at the group level

	Responsiveness	Clarity	Expressed Agreement
Perceived Ignoring	$-.76^{***}$	$.63^{***}$	$-.21^{ns}$
Perceived Inhibition Others	$.58^{***}$	$-.40^{*}$	$.16^{ns}$
Perceived Inhibition Self	$-.04^{ns}$	$.02^{ns}$	$-.00^{ns}$
Perceived Consensus	$.48^{**}$	$-.55^{***}$	$.28^{ns}$
Perceived Solidarity	$.64^{***}$	$-.51^{**}$	$.33^{ns}$

Note. ^{ns} $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$.
Reported are the repeated measures correlations calculated using the rmcorr package (version 0.3.0; Bakdash & Marusich, 2017). In order to link the content coding (group level) to the questionnaire data (participant level), we aggregated the latter over group members.

responsiveness was associated with feeling less ignored, perceiving more inhibition in others, and experiencing more consensus and solidarity. Clarity was also correlated rather strongly with participants' experience ratings (r ranging from |.40| to |.63|): clearer messages were associated with feeling more ignored, with perceiving less inhibition in others, and with experiencing less consensus and solidarity.

We also examined the relations between the social attributions and relational consequences. The participant level repeated measures correlations were all strong and in expected directions. Specifically, experienced consensus and solidarity related negatively to perceived ignoring (r = -.23, $p < .001$ and r = -.43, $p < .001$, respectively), and positively to perceived inhibited behavior (r = .43, $p < .001$ and r = .59, $p < .001$, respectively). Thus, participants who felt their interaction partners were ignoring them more and/or acted more disinhibited, experienced more disagreement with them and felt less closely connected to them. This is in line with our assumption that social misattributions are part of the process underlying the negative effects of unresponsiveness and clarity on consensus and solidarity.

In sum, the correlations suggest that the reduced responsiveness and enhanced clarity in the online reactions to dissent can partly explain why, when using that medium, participants felt more ignored and thought others were less inhibited, and why they (maybe via these social misattributions) experienced less consensus and solidarity within their group.

Discussion

In light of scientific and public concerns about the polarizing effects of online discussions, this article examined the social dynamics that are at play *within* these discussions. We specifically examined the diplomatic skills that interaction partners use to socially regulate dissent in text-based online discussions and how these affect their social relationships. This article built on previous evidence suggesting

that online discussions can be relatively unresponsive and clear, which might negatively affect social harmony (Roos, Postmes, & Koudenburg, 2020), by introducing grounds for conflict and testing misattributions as underlying process.

Social regulation: Behavior

Results show that discussion content did not significantly differ across media in terms of toxic disinhibition (i.e., hostile or aggressive expressions) or expressed disagreement (Hypothesis 1c). This is quite a striking finding, because we gave interaction partners a good reason to respond in unfriendly ways: there was a dissenting confederate who expressed strongly deviant views. These first results nuance the online disinhibition literature (Suler, 2004) because they suggest that participants in our study were not less inhibited online and remained motivated to avoid conflict.

However, we did observe considerable differences in the *style* of participants' reactions to the confederate's dissent: online these were less responsive (Hypothesis 1a) and more clear (Hypothesis 1b) than FtF. We propose that this is due to the intrinsic characteristics of text-based communication: a relative lack of synchronicity and more difficulty in conveying subtle social cues are features of written communication. In FtF discussions, people often embed their disagreements in highly responsive and ambiguous messages, which their interaction partners tend to interpret as signals of social interest and relational investment, and which thereby enable them to prevent conflict and maintain harmony (e.g., Bavelas et al., 1990; Beňuš et al., 2011; Pomerantz, 1984). Therefore, the relatively low levels of responsiveness and ambiguity in online discussions could have negative consequences for interaction partners' social perceptions and relational outcomes.

Social regulation: (Mis)Attributions

Indeed, we found that the unresponsiveness and clarity in text-based online discussions affected participants' social perceptions: participants thought their interaction partners ignored them more (Hypothesis 2a) and were more disinhibited (Hypothesis 2b). Thus, rather than recognizing that it results from the restrictions that the text-based medium imposes on behavior (external causes), it appears that participants misattribute at least part of the unresponsive and clear communication to their interaction partners' motivations (internal causes).

Notably, participants considered *themselves* to be neither more nor less disinhibited online. This suggests a fundamental attribution error may have been made (Jones & Harris, 1967): whereas participants seem to attribute their own behavior to the medium or do not even notice it, they attribute their partners' unresponsiveness and clarity to inattentiveness and self-centeredness. Thus, online messaging can promote *perceptions* of disinhibition in interaction partners that might not be related to any *actual* disinhibition, because communication is less interpersonally responsive and more forthright.

Online unresponsiveness often manifested itself in a disjointed conversation pattern where participants each followed their own individual line of reasoning. Uninterrupted by each other's expressions, people can, and apparently do, talk simultaneously in text-based online chats. Most likely as a consequence of this, participants felt ignored online; thinking their interaction partners were less interested in them and/or in what they had to say (see also Davis & Perkowitz, 1979; Koudenburg et al., 2017).

Whereas it is not surprising that unresponsiveness tends to be interpreted as a lack of social concern, it seems rather contra-intuitive that clarity can be too. This can be explained by considering that in FtF discussions, people are used to talk around disagreements to avert the threat these pose to social relationships (Bavelas et al., 1990; Pomerantz, 1984). Indeed, although the so-called "bald on

record" strategy is a direct, clear, and efficient way of expressing disagreement, it is considered not very polite (Brown & Levinson, 1987; Goldsmith & MacGeorge, 2000). Ambiguity in a context of disagreement might therefore be perceived as a sign of social concern: it shows to receivers that the sender is committed to maintaining their social relationship because he/she is exerting effort to stop their disagreement from escalating into conflict by wrapping it in tentative vagueness. When the sender leaves out this ambiguity, a bare statement of disagreement remains, leaving intact the associated threat to the social relationship. Receivers might think the sender does not value their relationship as he/she is not trying to avoid conflict.

Thus, in contrast to what could be inferred from the media richness literature (Daft & Lengel, 1986; Runions et al., 2013), our findings suggest that online interaction is *less* ambiguous in style. Moreover, our results show that ambiguity can be a good thing in the context of contentious discussions: it can blur differences of opinion and thereby serve to maintain social relationships.

Social regulation: Relational consequences

Because the greater unresponsiveness and clarity of online conversations can be misattributed, the social relationships of interaction partners could be affected. Indeed, results showed that participants perceived less consensus (Hypothesis 3a) and less solidarity (Hypothesis 3b) within their group after discussing online. These findings are in line with previous research showing that unresponsiveness feeds impressions of misunderstanding, dissent, and social rejection (Davis & Perkowitz, 1979; Koudenburg et al., 2013). The results also support the reasoning, based on the pragmatics literature, that ambiguity in contentious discussion can communicate the sender's intent to find consensus and his/her concern for the feelings of interaction partners.

Important to emphasize, however, is that even though participants perceived less consensus online, this was not reflected in any actual polarization in their privately held attitudes nor in their expressions of disagreement (Hypothesis 1c) in the discussions. This means that, as anticipated, there was a mismatch between real and perceived disagreement: while they did not disagree more, participants did *experience* more disagreement online. Again, there appears to be an attributional effect of misperception at play, which is not reflected in actual attitudes. We suggest that, like the misattributions of ignoring and disinhibited behavior, this misperception of polarization is due to the relatively unresponsive and clear style of text-based online discussions.

Attributions could partly drive the effect of style on social outcomes: receiving clearly phrased messages that do not respond to one's comments might make one feel ignored by one's interaction partners, who seem mainly concerned with acting on their own needs (i.e., disinhibited). This is likely to feed into the impression that one's interaction partners do not value one's standpoint and/or oneself as a person, and are not interested in reaching consensus and/or maintaining a good relationship. Note, however, that this reasoning is only based on correlational data and therefore remains conjectural.

In sum, the results suggest that even in the absence of actual polarization, text-based online discussions can give rise to *perceptions* of increased polarization, because interaction partners cannot resort to the everyday diplomatic skills routinely used to regulate verbal disagreements. It is noteworthy that most prior research has focused on actual polarization online, presumably because researchers assume this to be the problem. The current findings, however, show that merely perceiving polarization can damage social relationships: perceptions are clearly consequential (see also O'Sullivan & Flanagin, 2003).

More abstractly, the impression that one's interaction partners show relationally considerate behavior and are invested in maintaining a pleasant social relationship appears to be very important to conversational outcomes. We find that this impression is informed to a large extent by the style that one's interaction partners' reactions to disagreement take. Therefore, discussion *style* seems of decisive importance in steering conversational experiences (see also Koudenburg et al., 2017).

Limitations and future research

First, it is important to keep in mind the focus of this article: we studied groups of relative strangers handling disagreement on a controversial topic within a restricted time frame. We expect that the consequences of (a lack of) responsiveness and ambiguity are largest in this specific context as disagreement is unexpected and undesirable, and social relationships are budding and fragile. Whether our conclusions can be generalized to situations in which people know each other or discuss non-controversial topics, remains to be seen. Similarly, the present findings may not extend to contexts in which people actively seek out contention, such as a debating club.

Second, the external validity of the results should be tested. We studied young, mostly female, and highly educated Dutch students discussing with a disagreeing confederate via an instant text-based chat in the lab. These participants were probably motivated to keep the discussion pleasant. Accordingly, it would be wrong to generalize these results to online users that are intentionally being mean and disruptive (Hardaker, 2010). When the context would be less friendly, perceiving a sender as ignoring and disinhibited might provoke receivers to retaliate with uncivil posts, resulting in a vicious circle of aggravating conflict and polarization (Chen & Lu, 2017). Our results suggest that such vicious circles may be set in motion without harmful intent, but through the restrictions the medium imposed on everyday diplomatic skills. Nowadays, online communication takes many different forms: consider snapchat, reddit, or twitter. It would be interesting to extend the present findings to studying the dynamics of polarization in more large-scale online discussions.

A third limitation of this study is that we did not control for the effects of anonymity (or pseudonymity), which has been connected to the conflict proneness of online discussions (e.g., Suler, 2004). As noted before, the current research builds on a more explorative study by Roos, Postmes, & Koudenburg (2020). This previous study included an additional text-based online chat condition that was combined with a live video-stream (without audio). This non-anonymous condition produced very similar results to the pseudonymous condition with only text-based chat. Therefore, in the present research, we decided not to study the effects of visual identifiability any further. Moreover, the absence of an effect for order of conditions in the current study suggests that pseudonymity did not play a role here either: the half the sample that engaged in FtF discussion first were not pseudonymous to each other but did show the same effects on the dependent variables. In sum, although we cannot entirely rule out the influence of anonymity, it seems safe to conclude that it is not the sole or leading explanatory factor.

Fourth, one might suggest that the reduced solidarity online can be caused by the increased amount of time it takes to build social relationships when communicating via text-based online chat compared to FtF, due to the relatively longer time typing takes (Walther, Anderson, & Park, 1994). To account somewhat for the time factor, we gave participants a maximum time frame of 10 minutes within which they could collectively decide to end their discussion (which they did more often in the FtF condition). Further, to increase comparability, we only coded the participant speaking turns between the first and the third utterance of the confederate. The results showed that these parts of the discussions already differed significantly between conditions and that these differences correlated

strongly with participants' social perceptions. This suggests that the restricted volume of text-based online communication cannot explain the results.

Lastly, this study looked at the effects of (the lack of) responsiveness and ambiguity concurrently. This means that we cannot isolate the consequences of each factor. Future studies could try to dissect the two by manipulating one while holding the other constant. In addition, we cannot conclude anything about the role of valence in responsiveness: is responsiveness a pleasant experience in and of itself, no matter how positive (affirmative) or negative the reply, or is just being recognized not enough and should it imply a positive evaluation? These might be promising directions for further research.

Conclusion

In a world where concerns about the polarizing effects of online discussions are rising, there is a need for an accurate understanding of the diplomatic skills that are required to regulate and moderate social interactions to prevent excessive polarization. The results of the current study suggest that we need to move beyond some of the prevalent assumptions about text-based online interaction. Online disinhibition and polarization might not be due to a lack of social constraints causing individuals to act without inhibitions and to voice more disagreement, but due to a failure to correctly implement diplomatic skills, causing group members to *perceive* more disinhibition and polarization. The present results underscore the importance of social regulation when there is disagreement, and in particular the utility of everyday diplomatic skills, such as interpersonal responsiveness and substantive ambiguity, in discussing a contentious issue. On this basis, the practical recommendations following from this study are somewhat counter-intuitive. Specifically, interaction partners can promote harmony in their online discussions by giving more instant feedback that is non-substantive and by being more ambiguous in their contributions. In some sense, the irony of these results is that they show that for contentious issues, FtF communication is in some ways a superior medium precisely because in terms of accuracy and information content it is worse.

Acknowledgments

The authors thank Professor Tom Snijders for his statistical advice, Iris Koomen and Lieke Molenaar for acting as research confederates, and Lotte Hulleman for performing the content coding.

Data availability

The data underlying this paper will be shared on reasonable request to the corresponding author.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Notes

1. The study was approved by the Ethical Committee Psychology of the University of Groningen. Pre-participation informed consent was obtained from all participants.
2. As the inclusion of demographic variables, such as age or gender, as predictors in the models did not change the effects of condition on any of the dependent variables, the effects of these covariates will not be reported.
3. Intra-class correlation coefficients were calculated over all turn ratings by means of the ICC function in the IRR package (version 0.84.1, Gamer, Lemon, Fellows, & Singh, 2019).
4. We exploratively coded a couple of additional constructs but these are not reported as they turned out to be either unreliable or redundant. These are available from the corresponding author on reasonable request.
5. The questionnaire also included additional items for exploratory purposes, which are available on reasonable request to the corresponding author.
6. We report omega (hierarchical) with bias corrected and accelerated (1000) bootstraps. These were calculated with the ci.reliability function of the MBESS package (version 4.6.0, Kelley, 2019).
7. The scale used by Koudenburg et al. (2013) measures different aspects of shared cognition, such as understanding and being on the same wavelength. For this study we added a couple of items about perceived (dis)agreement to this scale.
8. Koudenburg et al. (2015) used 22 items to measure solidarity-related constructs. As participants had to fill out the same questionnaire twice (repeated measures), we chose to shorten this scale by selecting eight items.
9. We added four extra statements to the actual stimulus statements in the pre-discussion opinion measure to avoid priming participants.
10. We chose to perform a repeated measures correlation analysis as our design did not allow for a mediation analysis. Specifically, the hypothesized mediators (responsiveness and clarity) were measured at the group level (coded per conversation) while the outcomes (ignoring, disinhibition, solidarity and consensus) were measured at the individual participant level (questionnaire scores). There are not enough observations to use group aggregated questionnaire scores as dependent variables. However, relations among variables can be reliably inferred from correlation matrices and considering that the discussion content occurred before we questioned participants about their discussion experiences, the proposed causal path from content to experience seems the most plausible.

References

Alberici, A. I., & Milesi, P. (2018). Online discussion and the moral pathway to identity politicization and collective action. *Europe's Journal of Psychology*, *14*(1), 143–158. doi:10.5964/ejop.v14i1.1507

Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M.A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, *30*(1), 156–168. doi:10.1093/ijpor/edw022

Bakdash, J. Z., & Marusich, L. R. (2017). *Repeated measures correlation (version 0.3.0) [computer software and manual]*. Retrieved from https://cran.r-project.org/web/packages/rmcorr/rmcorr.pdf

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Fox, J. (2019). *Linear mixed-effects models using 'Eigen' and S4 (version 1.1-21) [computer software and manual]*. Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bavelas, J. B., Black, A., Chovil, N., & Mullet, J. (1990). *Equivocal communication*. Newbury Park, CA: Sage.

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001–3027. doi:10.1016/j.pragma.2011.05.011

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bolker, B. (2014). *Simulation-based power analysis for mixed models in lme4. [Computer software and manual]*. Retrieved from http://rpubs.com/bbolker/11703

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. doi:10.1037/0278-7393.22.6.1482

Brown, P., & Levinson, S. C. (1987). *Studies in interactional sociolinguistics, 4. Politeness: Some universals in language usage*. New York: Cambridge University Press.

Casale, S., Fiovaranti, G., & Caplan, S. (2015). Online disinhibition: Precursors and outcomes. *Journal of Media Psychology: Theories, Methods, and Applications*, *27*(4), 170–177. doi:10.1027/1864-1105/a000136

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. doi:10.1080/08838151.2016.1273922

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark-Gordon, C. V., Bowman, N. D., Goodboy, A. K., & Wright, A. (2019). Anonymity and online self-disclosure: A meta-analysis, *Communication Reports*, *32*(2), 98–111. doi:10.1080/08934215.2019.1607516

Coe, K., Kenski, K., & Rains, S.A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. doi:10.1111/jcom.12104

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NY: Lawrence Erlbaum Associates.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571. doi:10.1287/mnsc.32.5.554

Davis, D., & Perkowitz, W. T. (1979). Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology*, *37*(4), 534–550. doi:10.1037/0022-3514.37.4.534

Edwards, R., Bybee, B. T., Frost, J. K., Harvey, A. J., & Navarro, M. (2017). That's not what I meant: How misunderstanding is related to channel and perspective-taking. *Journal of Language and Social Psychology*, *36*(2), 188–210. doi:10.1177/0261927X16662968

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement (version 0.84.1) [computer software and manual]*. Retrieved from https://cran.r-project.org/web/packages/irr/irr.pdf

Geddes, D. (1992). Sex roles in management: The impact of varying power of speech style on union members' perception of satisfaction and effectiveness. *The Journal of Psychology: Interdisciplinary and Applied, 126*(6), 589–607. doi:10.1080/00223980.1992.10543390

Goldsmith, D. J., & MacGeorge, E. L. (2000). The impact of politeness and relationship on perceived quality of advice about a problem. *Human Communication Research*, *26*(2), 234–263. doi:10.1111/j.1468-2958.2000.tb00757.x

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan. (Eds.), *Syntax and semantics, Vol. 3, speech acts* (pp. 41–58). New York: Academic Press.

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, *6*(2), 215–242. doi:10.1515/jplr.2010.011

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. doi:10.20982/tqmp.08.1.p023

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of experimental social psychology*, *3*(1), 1–24. doi:10.1016/0022-1031(67)90034-0

Kelley, K. (2019). *MBESS (version 4.6.0) [computer software and manual]*. Retrieved from https://cran.r-project.org/web/packages/MBESS/MBESS.pdf

Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes*, *52*(1), 96–123. doi:10.1016/0749-5978(92)90047-B

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2013). Resounding silences: Subtle norm regulation in everyday interactions. *Social Psychology Quarterly*, *76*(3), 224–241. doi:10.1177/0190272513496794

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2017). Beyond content of conversation: The role of conversational form in the emergence and regulation of social structure. *Personality and Social Psychology Review*, *21*(1), 50–71. doi:10.1177/1088868315626022

Koudenburg, N., Postmes, T., Gordijn, E.H., & van Mourik Broekman, A. (2015). Uniform and Complementary Social Interaction: Distinct Pathways to Solidarity. *PLoS One*, *10*(6), e0129061. doi:10.1371/journal.pone.0129061

Krueger, J. (1998). On the perception of social consensus. *Advances in Experimental Social Psychology*, *30*, 163–240. doi:10.1016/S0065-2601(08)60384-6

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434–443. doi:10.1016/j.chb.2011.10.014

Lea, M., O'Shea, T., Fung, P., & Spears, R. (1992). 'Flaming' in computer-mediated communication: Observations, explanations, implications. In M. Lea, M. Lea (Eds.), *Contexts of computer-mediated communication* (pp. 89–112). London, England: Harvester Wheatsheaf.

Lenth, R., Singmann, H., Love, J., Buerkner P., & Herve, M. (2019). *Estimated marginal means, aka least-squares means (version 1.4.1) [computer software and manual]*. Retrieved from https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing 'flaming' and other problematic messages. *New Media & Society*, *5*(1), 69–94. doi:10.1177/1461444803005001908

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. doi:10.1177/1461444804041444

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 57–101). Cambridge, England: Cambridge University Press.

Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta- analysis. *Psychological Bulletin*, *123*(3), 238–259. doi:10.1037/0033- 2909.123.3.238

Reid, S. A., Keerie, N., & Palomares, N. A. (2003). Language, gender salience and social influence. *Journal of Language and Social Psychology*, *22*(2), 210–233. doi:10.1177/0261927X03022002004

Reis, H. T., & Clark, M. S. (2013). *Responsiveness*. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 400–423). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780195398694.013.0018

Roos, C. A., Postmes, T., & Koudenburg, N. (2020). The micro-dynamics of social regulation: Comparing the navigation of disagreement in text-based online and face-to-face communication. *Group Processes & Intergroup Relations*, *23*(6), 902–917. doi:10.1177/1368430220935989

Runions, K. C., Shapka, J. D., Dooley, J., & Modecki, K. (2013). Cyber-aggression and victimization and social information processing: Integrating the medium and the message. *Psychology of Violence*, *3*(1), 9–26. doi:10.1037/a0030511

Spears, R., Postmes, T., Lea, M., & Watt, S. (2001). A SIDE view of social influence In K. D. Williams and J. P. Forgas (Eds.), *Social influence: Direct and indirect processes* (pp. 331–350). Philadelphia, PA and Hove, England: Psychology Press.

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, *7*(3), 321–326. doi: 10.1089/1094931041291295

Terry, C., & Cain, J. (2016). The emerging issue of digital empathy. *American Journal of Pharmaceutical Education*, *80*(4), 58. doi:10.5688/ajpe80458.

Voggeser, B. J., Singh, R. K., & Göritz, A. S. (2018). Self-control in online discussions: Disinhibited online behavior as a failure to recognize social cues. *Frontiers in Psychology*, *8*, 2372. doi: 10.3389/fpsyg.2017.02372

Walker, H. M., & Lev, J. (1953). Analysis of variance with two or more variables of classification. In H. M. Walker & J. Lev (Eds.), *Statistical inference* (pp. 348–386). New York: Henry Holt and Company.

Walther, J.B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, *19*(1), 52–90. doi:10.1177/009365092019001003.

Walther, J. B., Anderson, J. F., & Park, D. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and anti-social communication. *Communication Research*, *21*, 460–487. doi:10.1177/009365094021004002

Walther, J. B., Loh, T., & Granka, L. (2005). Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of Language and Social Psychology*, *24*(1), 36–65. doi:10.1177/0261927X04273036

Weber Shandwick, Powell Tate, & KRC Research (2013). *Civility in America 2013*. Retrieved from http://www.webershandwick.com/uploads/news/files/Civility_in_America_2013_Exec_Summary.pdf